

WHAT IS CLAIMED IS:

1. A statistical dialog system, comprising:
  - a speech understanding mechanism for determining the literal meaning of input speech data;
  - a dialog semantics learning mechanism for establishing semantic models based on annotated dialog training data, said annotated dialog training data associating literal meaning of input speech data with one or more semantic meanings of the input speech data; and
  - a statistical dialog manager for interpreting one semantic meaning of the input speech data based on both the literal meaning of the input speech data and corresponding semantic models that are associated with the literal meaning of the input speech data.
2. The system according to claim 2, wherein the speech understanding mechanism comprises:
  - a speech recognition mechanism for recognizing a word sequence from the input speech data based on at least one acoustic models; and
  - a language understanding mechanism for understanding the literal meaning of the word sequence based on a language model.
3. The system according to claim 1, further comprising a responding mechanism for generating at least one response to the input speech data based on the semantic meaning of the input speech data.
4. The system according to claim 3, wherein said responding mechanism includes:

a voice response mechanism for generating a voice response to the input speech data based on the semantic meaning of the input speech data; and

an action response mechanism for activating an action corresponding to the semantic meaning of the input speech data.

5. The system according to claim 4, wherein said voice response mechanism comprises:

a language response generation mechanism for generating a language response to the input speech according to the semantic meaning of the input speech data; and

a text to speech engine for synthesizing the voice of the language response to generate said voice response.

6. A dialog semantics learning mechanism, comprising:

an annotated dialog training data storage for storing annotated dialog training data that is either annotated off-line or fed back by a statistical dialog manager during on-line dialog sessions; and

a dialog semantic modeling mechanism for establishing semantic models of dialogs based on the annotated dialog training data.

7. The mechanism according to claim 6, further comprising a semantic model storage for storing the semantic models established by the dialog semantic modeling mechanism based on annotated dialog training data.

8. A system, comprising:

a semantic model retrieval mechanism for retrieving, from a semantic model storage, semantic models that are associated with a literal meaning of input speech data; and

a dialog semantic understanding mechanism for interpreting, during a dialog session, the semantic meaning of the input speech data according to said semantic models and said environmental status.

9. The system according to claim 8, further comprising:

an environmental status access mechanism for accessing environmental status that affects the interpretation of the semantic meaning of the input speech data, said environmental status being used, together with the semantic models, by the dialog semantic understanding mechanism to interpret the semantic meaning of the input speech data; and

a dialog data annotation mechanism for annotating the relationship between said literal meaning of the input speech data and the semantic meaning of the input speech data based on the dialog session to generate feedback dialog data.

10. A method, comprising:

receiving, by a statistical dialog system, input speech data;

determining, by a speech understanding mechanism in the statistical dialog system, the literal meaning of the input speech data;

retrieving at least one semantic model associated with the literal meaning of the input speech data, said at least one semantic model associating the literal meaning with at least one semantic meaning of the input speech data;

interpreting, by a statistical dialogue manager in the statistical dialogue system, the semantic meaning of the input speech data based on the literal meaning of the input speech data and the at least one semantic model; and

generating a response to the input speech data based on the semantic meaning of the input speech data.

11. The method according to claim 10, wherein said determining the literal meaning comprises:

recognizing, by a speech recognition mechanism, a word sequence from the input speech data based on at least one acoustic model; and

generating, by a language understanding mechanism, a literal meaning of the input speech data from the word sequence based on a language model.

12. The method according to claim 10, wherein said generating a response includes at least one of:

generating, by a voice response mechanism, a voice response to the input speech data based on the semantic meaning of the input speech data; and

generating, by an action response mechanism, an action response to the input speech data according to the semantic meaning of the input speech data.

13. The method according to claim 12, wherein said generating a voice response comprises:

producing, by a language response generation mechanism, a language response according to the semantic meaning of the input speech data; and

synthesizing, by a text to speech engine, the voice of said language response to generate said voice response.

14. A method for dialog semantic learning, comprising:

receiving annotated dialog training data that associates a literal meaning of input speech data with at least one semantic meaning of the input speech data; and

training a semantic model corresponding to the literal meaning of the input speech data based on the annotated dialog training data.

15. The method according to claim 14, further comprising:

storing the semantic model in a semantic model storage.

16. A method for a statistical dialog manager, comprising:

receiving, from a speech understanding mechanism, a literal meaning corresponding to input speech data;

retrieving, from a semantic model storage, at least one semantic model associated with the literal meaning of the input speech data; and

interpreting, by a dialog semantic understanding mechanism, the semantic meaning of the input speech data based on the literal meaning of the input speech data and the at least one semantic model.

17. The method according to claim 16, wherein said interpreting the semantic meaning comprises:

determining at least one semantic meaning of the input speech data according to the literal meaning and the at least one semantic model; and

confirming, based on the at least one semantic meaning of the input speech data, the semantic meaning associated with the literal meaning in a dialog session.

18. The method according to claim 17, further comprising:

accessing environmental status that affects the interpretation of the semantic meaning of the input speech data, said environmental status being used, together with the at least one semantic model, by said interpreting to generate the semantic meaning of the input speech data; and

annotating, by a dialog data annotation mechanism, the relationship between said literal meaning of the input speech data and the semantic meaning of the input speech data, confirmed during the dialog session, to generate feedback dialog data.

19. A computer-readable medium encoded with a program, said program comprising:

receiving, by a statistical dialog system, input speech data;

determining, by a speech understanding mechanism in the statistical dialog system, the literal meaning of the input speech data;

retrieving at least one semantic model associated with the literal meaning of the input speech data, said at least one semantic model associating the literal meaning with at least one semantic meaning of the input speech data;

interpreting, by a statistical dialogue manager in the statistical dialogue system, the semantic meaning of the input speech data based on the literal meaning of the input speech data and the at least one semantic model; and

generating a response to the input speech data based on the semantic meaning of the input speech data.

20. The medium according to claim 19, wherein said determining the literal meaning comprises:

recognizing, by a speech recognition mechanism, a word sequence from the input speech data based on at least one acoustic model; and

generating, by a language understanding mechanism, a literal meaning of the input speech data from the word sequence based on a language model.

21. The medium according to claim 19, wherein said generating a response includes at least one of:

generating, by a voice response mechanism, a voice response to the input speech data based on the semantic meaning of the input speech data; and

generating, by an action response mechanism, an action response to the input speech data according to the semantic meaning of the input speech data.

22. The medium according to claim 21, wherein said generating a voice response comprises:

producing, by a language response generation mechanism, a language response according to the semantic meaning of the input speech data; and

synthesizing, by a text to speech engine, the voice of said language response to generate said voice response.

23. A computer-readable medium encoded with a program for dialog semantic learning, said program comprising:

receiving annotated dialog training data that associates a literal meaning of input speech data with at least one semantic meaning of the input speech data; and

training a semantic model corresponding to the literal meaning of the input speech data based on the annotated dialog training data.

24. The medium according to claim 23, said program further comprising:  
storing the semantic model in a semantic model storage.

25. A computer-readable medium encoded with a program for a statistical dialog manager, said program comprising:

receiving, from a speech understanding mechanism, a literal meaning corresponding to input speech data;

retrieving, from a semantic model storage, at least one semantic model associated with the literal meaning of the input speech data; and



interpreting, by a dialog semantic understanding mechanism, the semantic meaning of the input speech data based on the literal meaning of the input speech data and the at least one semantic model.

26. The medium according to claim 25, wherein said interpreting the semantic meaning comprises:

determining at least one semantic meaning of the input speech data according to the literal meaning and the at least one semantic model; and

confirming, based on the at least one semantic meaning of the input speech data, the semantic meaning associated with the literal meaning in a dialog session.

27. The medium according to claim 26, said program further comprising:

accessing environmental status that affects the interpretation of the semantic meaning of the input speech data, said environmental status being used, together with the at least one semantic model, by said interpreting to generate the semantic meaning of the input speech data; and

annotating, by a dialog data annotation mechanism, the relationship between said literal meaning of the input speech data and the semantic meaning of the input speech data, confirmed during the dialog session, to generate feedback dialog data.